Comments on simulation model 22Aug21

**Experiments**

**Sim 1 (purely spatial) with the goal to estimate l1**

X = l0 + l1L + e

e = eR + eF

eR ~ Normal N(0, sS) with S = f(D,q) [exclude the nugget] where f() is the tapered covariance function

eF ~ fixed 2D sine function (formerly R, but since it is fixed, I changed the name)

L = 0 or 1 with a 6x6 grid (36 boxes)

a. effect of scale of random spatial autocorrelation in e, q

e = eR (i.e., eF = 0)

q = 0, 0.05, 0.25

l0 = 0

l1 = 0.2

give table of est l1 - true l1 from 200 simulations

b. non-Gaussian errors

same as 1a but with a t3 distribution of eR. Then for the case with l1 = 0 and q = 0.05, do 1000 simulations and report the rejection rate (check for type 1 errors). This is particularly interesting for the case of non-Gaussian errors.

c. effect of scale of fixed spatial autocorrelation in eF

(The reason this is interesting is because kriging approach use fixed rather than random spatial components in the models)

e = eR + eF

q = 0, se small enough that eF dominates

Tx = Ty picked to give 22, 32, 42 peaks. (To make this a challenge, could you make it so when there are 32 peaks, they correspond to the centers of the boxes with L = 1? )

l0 = 0

l1 = 0.2

give table of est l1 - true l1 from 200 simulations (or add to previous table for Sim1a, b)

[d. map size - maybe we should just drop this, since it takes a long time.]

**Sim 2 (spatiotemporal) with tests of time trends**

X(t) = l0 + (b0 + b1L)t + e(t)

e(t) = r e(t-1) + d(t)

d(t) ~ Normal N(0, sS) with S = f(D,q) [exclude the nugget] where f() is the tapered covariance function

L = 0 or 1 with a 6x6 grid (36 boxes)

q = 0, 0.05, 0.25

l0 , b0 = 0

r = 0.4

sd = 1/(1-r2)

b1 = 1/30

give table of est b1 - true b1 from 200 simulations

**Sim 3 (spatiotemporal) with test of a spatiotemporal independent variable**

X(t) = l0 + gW(t) + e(t)

e(t) = r e(t-1) + d(t)

W(t) = rW W(t-1) + dW(t)

dW(t) ~ Normal N(0, sWS) with S = f(D,qW) [exclude the nugget] where f() is the tapered covariance

q = 0.25

qW = 0, 0.25

l0 = 0

r = 0.4

rW = 0, 0.4

sd = 1/(1-r2)

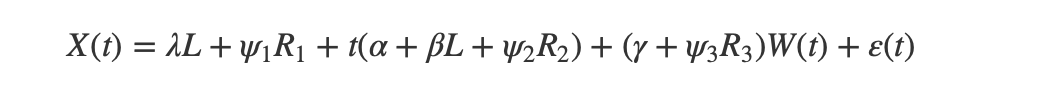
sW = 1/(1-rW2)

g = 1

give table of est g - true g from 200 simulations

Comments on previous versions

Original model



New simplified model

W(t) = r1 W(t-1) + d1(t)

WU(t) = rU WU(t-1) + dU(t)

U(t) = hW(t) + WU(t)

X(t) = l0 + l1L + (b0 + b1L)t + gW(t) + sUU(t) + e(t)

where W(t) is measured and U(t) isn't. Therefore, U(t) serves as the source of spatial and temporal autocorrelation in the random variation. Note that U(t) can contain W(t) so that there is correlation between the errors and W(t) to make estimation of g hard.

e ~ Normal N(0, S) with S = f(D,theta) [exclude the nugget] where f() is the tapered covariance function

d1 ~ R + N(0, small sd)

dU ~ Normal scaled so the variance of W(t) is 1

I took out R from X(t) (the 2D sine function) because I'm really trying to simplify things. I personally really like having in R, but this gives 3 different ways in which spatial variation is produced.

Granularity fixed at 6

For time-series simulations, have a burn-in time of 10 years.

t-distribution with 3 df

**Sim1a:** Test the impact of map size on PARTs ability to estimate only spatial effects.

a. Map widths of 104, 144, 200, and 280 pixels

b. Beta0 = beta1 = gamma = 0

c. Random term d2 ~ Normal

d. Set g = h = f = 0. I'm not sure the base values for the other parameters, but use the ones you have been using if they give reasonable simulations. Well, maybe increase q to 0.1 to give more spatial autocorrelation.

e. Set l1 = 0 and l1 > 0 in two sets of simulations; for the l1 > 0, it would be nice to pick l1 so that the rejection rate for a significance level of 0.05 was about 30% for the smallest maps, so this simulation would show the effect of map size on power.

f. Plot simulation example of U

g. Plot est l1 - sim l1 and rejection rates if you can do 500 simulations)

Sim1b: Test the impact of non-gaussian errors on estimating only spatial effects.

a. Do a study like sim1a with 104 pixel maps, but with random term d2 ~ T3

b. Plot est l1 - sim l1; only do rejection rates if you can do 500 simulations (i.e., these aren't as important)

c. plot comparison of l1 among sim 1a (104 px) and sim 1b (i.e., compare gaussian vs T)

Sim1c: Test the impact of introducing random spatial variation to the spatial parameter.

\* (this replaces old sim2 that looks a granularity, doing so without time; it is also similar to old sim 3 in looking at grain)

a. Do a study like sim1a with 104 pixel maps, but f > 0 (R included), with Tx = Ty picked to give 1, 22, 32, ..., 82 peaks. At 82 peaks, estimates of l1 should be bad, since this is the same scale as the granularity.

b. Plot est l1 - sim l1 and rejection rates (only if you can do 500 simulations), including comparison from when f = 0 from sim 1a

Sim2a: Test the effect of temporal autocorrelation on estimates of **intercept** time trends

a. 30 time points

b. g = h = 0

c. lambda = gamma = 0

d. b0 = 0 and b0 = 1/30; b1 = 0

e. To make things hard, use r2 = 0 and r2 = 0.6 (i.e., temporally autocorrelated U)

f. Plot est b0 - sim b0 but not rejection rates

Sim2b: Test the effect of temporal autocorrelation on estimates of time trends **interaction effects**.

a. 30 time points

b. g = h = 0

c. lambda = gamma = 0 (like 2a)

d. b0 = 0; b1 = 1/30

f. To make things hard, use r2 = 0 and r2 = 0.6 (i.e., temporally autocorrelated U)

g. Plot est b1 - sim b1 but not rejection rates

Sim3: Test the effects of a nugget on estimates of climate variable effects (gamma)

a. Simulate cases g = 0 and g > 0 (maybe g = 0.1 like you used before)

b. r1 = r2 = 0.6, b0 = b1 = 0

c. lambda = 0??

c. Simulate two cases h = 0 and h > 1

d. Plot est g - sim g but not rejection rates

Total time estimates (not accounting the burn-in period):

* If we simulate 500 datasets for each case, it will take a total of 31 days on 4 cores:
  + Sim 1: 498 hrs = 20.75 days
    - 1a: 310 hrs
    - 1b: 38 hrs
    - 1c: 150 hrs
  + Sim 2: 113 hours = 4.7 days
    - 2a: 75 hrs
    - 2b: 38 hours
  + Sim 3: 150 hours = 6.25 days
* If we reduce this to 200 per case, it might take roughly 31\*0.4 = 12.4 days and calculations of rejection rates would be questionable.